

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 72 (2015) 461 – 468

Procedia
 Computer Science

The Third Information Systems International Conference

Slice and Connect: Tri-Dimensional Process Discovery with Case Study of Port Logistics Process

Iq Reviessay Pulshashi^a, Hyerim Bae^{a*}, Riska Asriana Sutrisnowati^b,
Bernardo Nugroho Yahya^c, Sanghyuck Park^a

^aDepartment of Industrial Engineering, Pusan National University, Busan 609-735, South Korea^bDepartment of Big Data, Pusan National University, Busan 609-735, South Korea^cDepartment of Industrial Engineering, Hankuk University of Foreign Studies, Yongin 449-791, South Korea

Abstract

One problem of process discovery is the duplicated task, a task that appears multiple times in one process model. This paper presents “Slice and Connect,” an automated tri-dimensional process discovery technique considering duplicated task. It adds the notion of pool as a group of tasks that belongs to a single entity (e.g. department, organization, country). The algorithm entails two steps: first, the event log is sliced according to the pool attribute, and then the pool dependency network (PDN) is discovered. Second, the inter-pool dependency is calculated to connect the PDNs, the result of which is a process model represented as an integrated dependency network (IDN). Duplicated task is only allowed within a different pool to clearly separate the activity context in IDN. The complexity of the process model thereby is increased; therefore, the comprehensibility of the process model is also increased.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015)

Keywords: process mining; process discovery; multi-dimensional process model; BPMN; port logistics

1. Introduction

Process mining is a process-centric methodology for extraction of knowledge from an event log, a log that contains historical data on activities generated by either humans or machines [1]. There are three types of process mining: 1) process discovery aims to discover a process model automatically from an event log, 2) conformance checking aims to test the conformance of a log with a model or a model with a log and 3) enhancement aims to enrich the current model to produce a new process model (Figure 1).

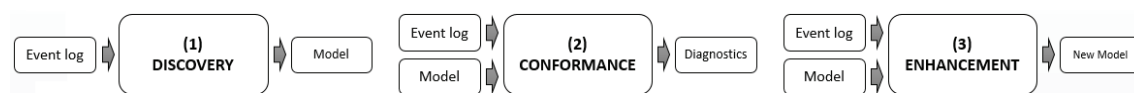


Figure 1. Three basic types of process mining in terms of input and output. Adapted from [2].

Process discovery have been an important research issue in process mining. Discovering such a “fit” process model is quite challenging [1]. First, there is representational bias, a situation wherein the discovered model cannot be visualized properly. Also difficulties dealing with complex event logs while balancing between quality criteria such as fitness, simplicity, precision and generalization for a discovered model. To improve process discovery, several approaches have been introduced such as alpha miner [3], heuristics miner [4] and fuzzy miner [5]. Alpha miner uses petri net to represent a process model however the algorithm is not robust since it was unable to deal with noise. Heuristic miner and fuzzy miner, attempt to cover the alpha miner limitations using some quality measures, which parameters are used to adjust a process model granularity. Regardless of the algorithm, process discovery always deals with the concept of causality to mine the control-flow. Therefore, the control-flow result, which is to say the process model, is often represented as a (activity) dependency network (Figure 2).

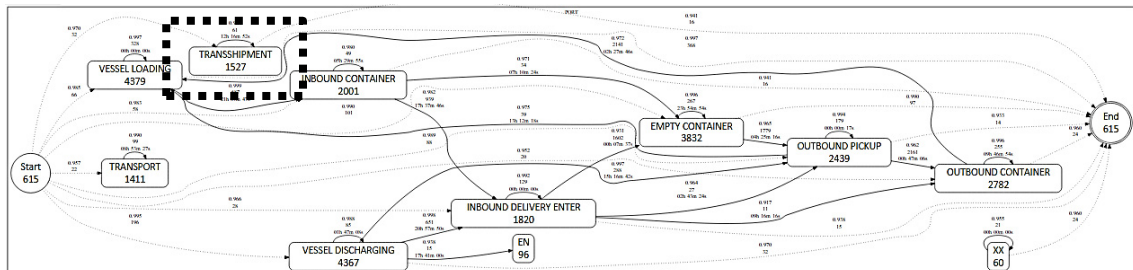


Figure 2. Port logistics process discovered using Heuristic Miner

This study aims to develop a new approach to handle, in particular, the existence of duplicated task. This paper introduces a methodology called Slice and Connect to discover a tri-dimensional process model considering duplicated task. The algorithm entails two steps: first, a process cube (Section 2.1) is sliced in a single dimension called a pool dimension (Section 3.1), and bi-dimensional process discovery is applied for each pool (Section 2.2); second, the activity is connected between pools (Section 3.1). The result is represented as a dependency network with pools. The methodology was validated using a case study of a port logistics process (Section 4). Last, we provide a discussion and anticipate future work related to tri-dimensional process discovery (Section 5).

2. Preliminaries

Originator, an agent of activity, is another important perspective or dimension in process mining. This dimension focuses on the relations among originators. The social network algorithm [6] uses several relationship measures such as handover work, working together and similar task. Later, several algorithms are introduced considering multiple dimensions, such as the Slice, Mine and Dice algorithm [7], process cube [8] and multi-dimensional time-gap analysis [9]. Additionally, bi-dimensional process discovery [10] is introduced considering two dimensions, activity and originator. Bi-dimensional process discovery, also known as BPMN (Business Process Model Notation) Miner, uses BPMN as a representation, and then derives the concept of a pool, a group of tasks that are assigned to a particular respondent. The algorithm consist of two steps: first, heuristic miner is used to mine the dependency network; second, a social network is calculated from the originator dimension, and a pool is mined by grouping all activities with the same originator, or, if the activity involves multiple originators, the pool is combined; finally, the particular process model is converted to BPMN.

One of the challenging on mining the control flow is the existence duplicated task. A duplicated task is regarded as a task that may appear multiple times in a process model. A duplicated task may exist as a part of noise, a loop or a cause of structuring business process (Figure 3). This situation eventually occurs when a particular task is used more than once in different context such as business entities within a process model. For example, in a port logistics process, there is a “customs” activity on the origin side and another “customs” activity on the destination side. In fact, this incurs ambiguity problems and increases the complexity of process model discovery. To tackle the problem of duplicated tasks, several process discovery algorithms have been introduced, including the region-based algorithm [13], the genetic algorithm [14], artificial negative events [15], and local search [16]. Unfortunately, since those algorithms consider only the activity dimension, it is not easy to find different entities in a process model, because there is the possibility that duplicated tasks exist in some or any of the process model sequences.

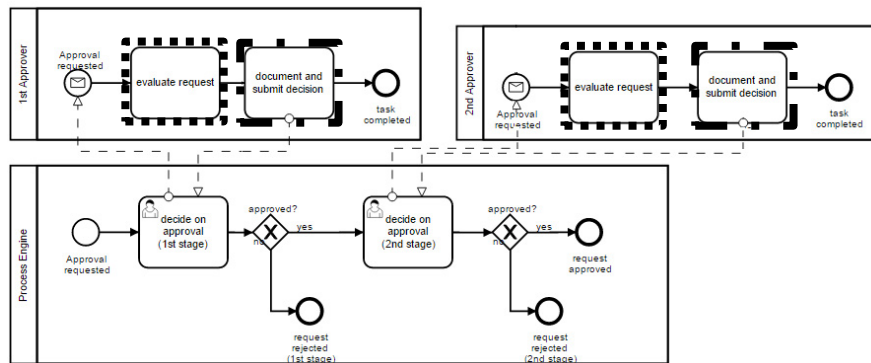


Figure 3. BPMN Model with Duplicated Task. Adapted from [12]

Suppose that we have an event log that involves several tasks (activities) by several roles (originators) in several locations (pools). The work name and role name is shared between locations. Using a heuristic miner, we can obtain the overall process model represented as a dependency network, and then we conduct a social network analysis to acquire the overall network model of the originators. Bi-dimensional process discovery will improve the comprehensibility of the process by adding the concept of a pool. However, it allows mapping of only one activity into single pool, or the pool will be merged. Since the work name and role name are shared, it cannot tackle the problem of duplicated tasks among them, and becomes only one large pool in one process model (Figure 4).

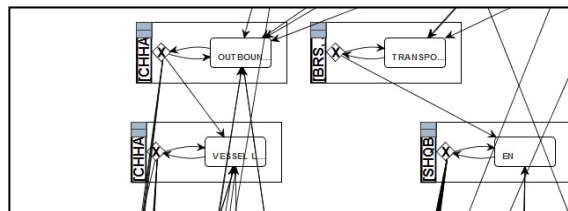


Figure 4. Problem with current BPMN Miner: includes concept of pool but does not allow any duplicated task in single process model

2.1. From Event Log to Process Cube

We define the following terms and notations. An event log L is defined as a multi set of traces. Each traces set $\sigma \in L$ is a finite sequence of events with σ_i the event at position i in trace σ . Each event has

properties that contain dimensions and values denoted by $\#_{dim}(\sigma_i) = value$. In our case, it is necessary for each event to have at least five dimensions: $\#_{case}$, $\#_{activity}$, $\#_{timestamp}$, $\#_{originator}$ and $\#_{pool}$. Then, the structure of the log can be represented as a tri-dimensional cube with respect to the originator, timestamp and pool dimensions.

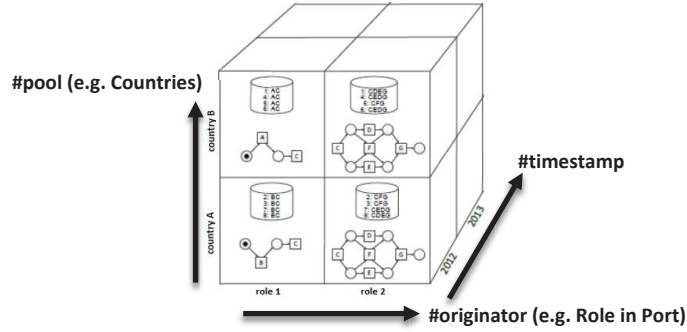


Figure 5. Event log and its corresponding process cube. Adapted from [8].

2.2. From Heuristic Miner to BPMN Miner

The heuristic miner algorithm uses direct the succession frequency and dependency score between two activities. It allows filtering of a certain transition that is lower than our threshold for dealing with noise, an infrequent behavior for an event log. First, the algorithm will calculate the direct succession frequency between activities as well as the dependency between them. Second, the split join can be determined as to whether it is “AND”, “OR” or “XOR”. The result is represented as an activity dependency network. BPMN Miner combines this score with the social network score to mine the activity pool in one process model, which is represented as swim lane on BPMN.

$|a \Rightarrow_L b|$ is the value of the dependency relation between a and b:

$$|a \Rightarrow_L b| = \begin{cases} \frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\ \frac{|a >_L b|}{|a >_L b| + 1} & \text{if } a = b \end{cases}$$

Figure 6. Dependency measure between two activities used by heuristic miner and later by BPMN miner. Adapted from [4]

3. Slice and Connect: Tri-dimensional Process Discovery

3.1. Slice

Our algorithm actually is quite straightforward. First we perform a one-direction slice on a process cube in the pool dimension. For each pool, we apply heuristic miner or BPMN miner to mine the process with certain positive observations and dependency thresholds. Then, the slice step is completed after the discovered process of each pool is represented as a dependency network called pool dependency network (PDN). PDN can be one-dimensional (heuristic miner, activity dimension) or two-dimensional (BPMN miner, activity and originator dimension) depends on algorithm we choose.

3.2. Connect

In this step, first, we calculate the dependency between pools: the inter-pool positive observations and the inter-pool dependency threshold. Then, if there is such dependency between two pools, inter-pool activity transitions are discovered and dependency are calculated then filtered using the inter-pool activity positive observation threshold and inter-pool activity dependency threshold. Finally, we represent the pool as a subgraph and connect it into single dependency network called integrated dependency network (IDN). This approach allows a duplicated task in one process model but only in a specific context (pool).

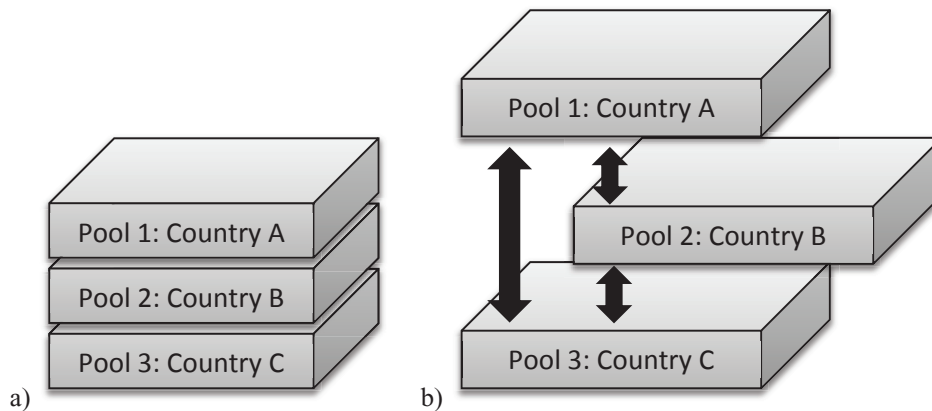


Figure 7. Slice and Connect: a) Slice process cube in pool dimension producing pool dependency network (PDN), b) Connect pool based on inter-pool dependency to create integrated dependency network (IDN)

4. Case Study

4.1. Port Logistics Process

To validate our approach, we used an event log from port logistics as a case study. In the case of one container, it is can be used by several logistic companies and visits several countries during the logistics process. The processes between ports are independent of each other, and so the problem is to discover an overall process model for later performance analysis. So, the event log contains a container as the case, a booking company as the originator, a country as the pool, and a logistic activity and corresponding timestamp (Figure 8).

CASE	ACTIVITY	TIMESTAMP	ORIGINATOR	POOL
CNT000001	VESSEL LOADING	2014-01-03 08:13:00	ROLE 1	COUNTRY A
CNT000001	VESSEL DISCHARGING	2014-01-25 08:13:00	ROLE 1	COUNTRY B
CNT000001	TRANSHIPMENT	2014-01-25 08:13:00	ROLE 1	COUNTRY B
CNT000001	VESSEL LOADING	2014-01-28 12:46:00	ROLE 1	COUNTRY B
CNT000001	VESSEL DISCHARGING	2014-02-04 22:00:00	ROLE 1	COUNTRY B
CNT000001	INBOUND CONTAINER	2014-02-04 22:00:00	ROLE 2	COUNTRY B
CNT000001	TRANSPORT	2014-02-05 23:10:00	ROLE 2	COUNTRY B
CNT000001	INBOUND CONTAINER	2014-02-05 23:10:00	ROLE 2	COUNTRY B
CNT000001	INBOUND DELIVERY ENTER	2014-02-24 23:00:00	ROLE 2	COUNTRY B

Figure 8. Event log of port logistic process

4.2. Process Discovery Result

We conducted an experiment using the open-source process mining tool ProM 6.4. To contrast our approach with others, we tried to mine the event log using heuristic miner and the BPMN miner package. Using heuristic miner, we could obtain the overall process of port logistics and represent it as a dependency network (Figure 2). In this case, it was clear that we could not discover the originator directly, or the process model would become too complex. It was also clear that the current BPMN miner is not suitable for mining such a duplicated tasks, since it will always map activities to a single pool or merge pools if it involves the same activity (Figure 4). Therefore, we modelled our approach in two steps: the first, slice step slices the process cube in the pool dimension and mined the PDN (Figure 9); the second, connect step connects those pools into single IDN for process model representation (Figure 10).

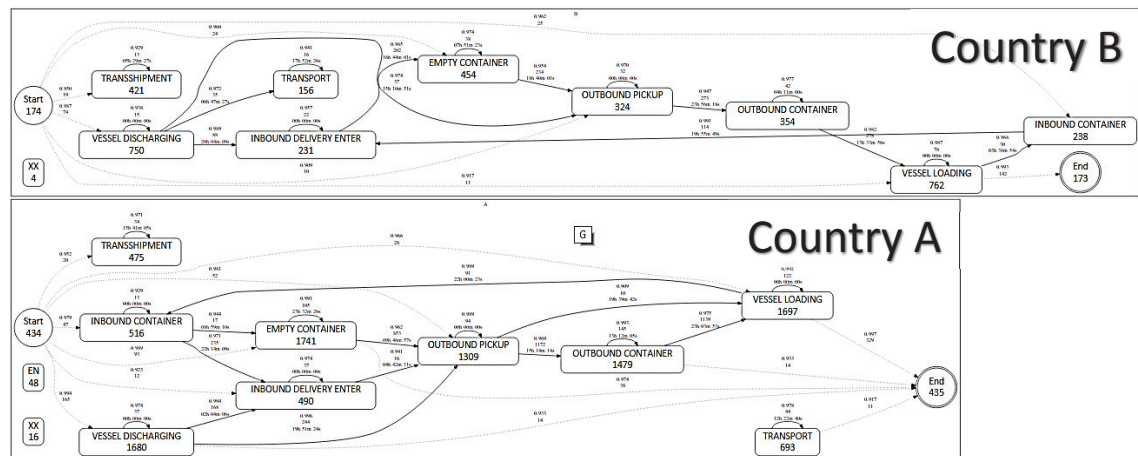


Figure 9. Pool dependency networks (PDN) of Country A and Country B discovered in slice step

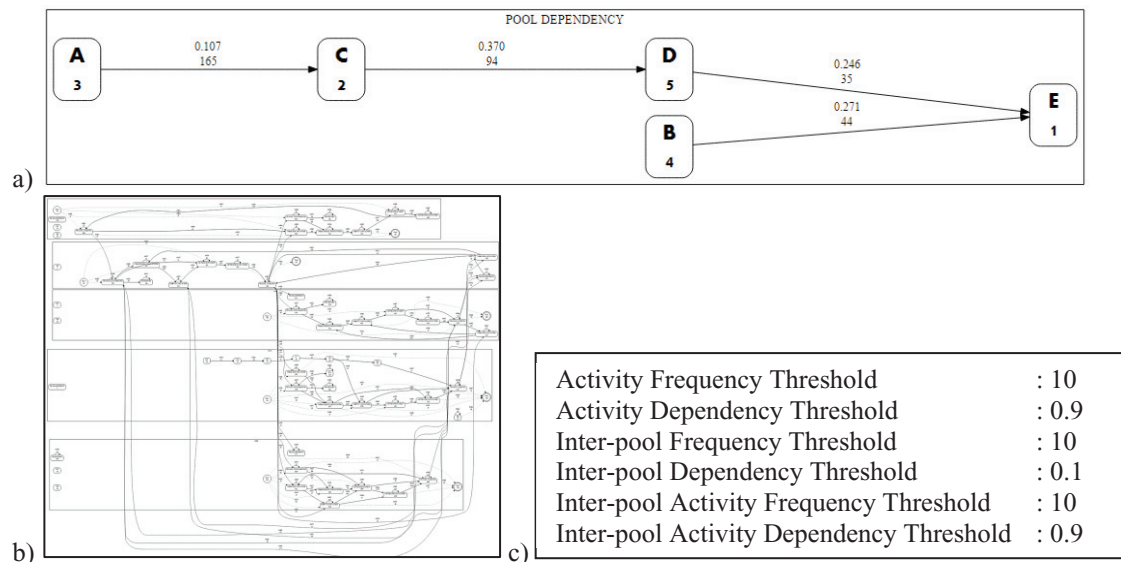


Figure 10. a) Inter-pool dependency network, b) bird view of integrated dependency network (IDN) discovered in connect step and c) Algorithm parameters

4.3. Use Case Scenario

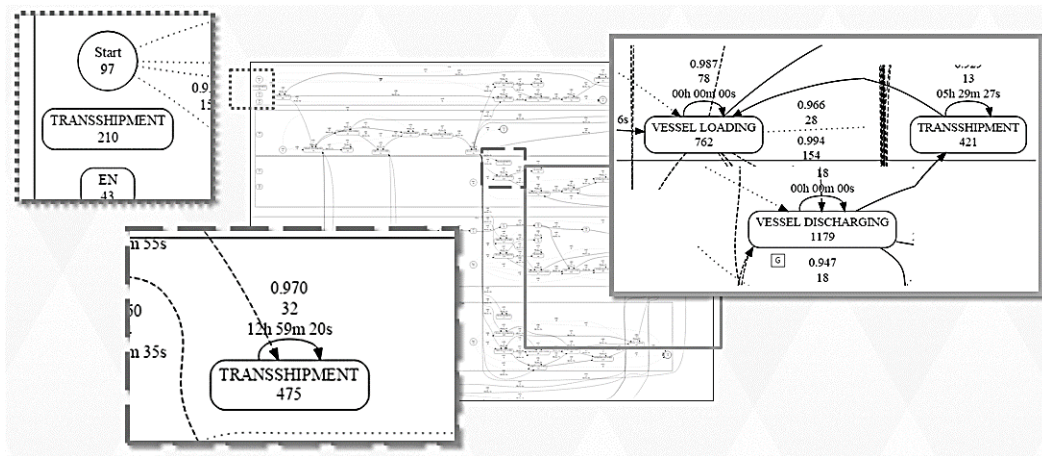


Figure 11. Comprehensive analysis can be explained from final integrated dependency network. Case study transshipment activity of port logistics

The necessity of mining a complete network is the need to find a different use context for the same activity. For example, a transshipment, a process of shipping a container to an intermediate destination before the final destination can be described more comprehensively. Using heuristic miner, we learned that transshipment is a process independent of the main process. Then, using our approach, we could explain more comprehensively based on the different pools ([13]); however, later, the same approach was used to mine the pool dependency networks. First, we could explain the dependency between the pools (country locations) using pool dependency networks. Second, we could explain the missing link in the transshipment process by considering the inter-pool dependency; we found that transshipment obtains input from vessels discharging in another country and outputs to vessels loading in another country.

5. Discussion

5.1. Conclusion

The necessity of discovering a duplicated task is increased when we can clearly define the different context and also when comprehensive analysis is required. By explicitly adding one more dimension, a pool dimension in an event log, we can clearly separate the context for each duplicated task. However, using the originator dimension, we can apply bi-dimensional process discovery to discover processes within one pool. Dependency for each pool also is considered; then, comprehensive analysis of a particular activity can easily be carried out from the interaction between each pool. Later, it can be used as base model for more complex analysis such as Bayesian networks [18] or multi-dimensional time-gap analysis [9].

5.2. Future Work

The approach was inspired by BPMN 2.0, a popular modelling language for business processes. For the future work, firstly we will enhance the representation as a complete BPMN model. Since this will also increase the complexity of the process model, some reduction will have to be performed. Then,

suitable conformance checking and performance analysis can be developed for handling such a process model in BPMN.

Acknowledgements

This study was supported by BK21 plus through a National Research Foundation (NRF) grant funded by the Korean Ministry of Education, by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2012R1A1A2008335), and by the IT R&D program of MOTIE/KEIT (No. 10054514): Automation of optimal process design for flexible and customizable manufacturing with on-demand order specification.

References

- [1] Van der Aalst WMP. Process Mining: Discovery, Conformance and Enhancement of Business Process. 20011. Springer.
- [2] IEEE Task Force of Process Mining. Process Mining Manifesto. http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:process_mining_manifesto. Accessed August 1, 2015.
- [3] Van der Aalst WMP, Weijters AJMM, Maruster L. Workflow Mining: Discovering process models from event logs. *IEEE IKDE* 2004; 16-9:1128-1142.
- [4] Weijters AJMM, Ribeiro JTS. Flexible Heuristics Miner (FHM). CIDM 2010.
- [5] Günther CW, van der Aalst WMP. Fuzzy mining - adaptive process simplification based on multi-perspective metrics. In: Alonso G, Dadam P, Rosemann M, editors. BPM 2007. *LCNS*; 4714: 328-343. Springer.
- [6] Van der Aalst WMP, Song M. Mining Social Networks: Uncovering Interaction Patterns in Business Processes. BPM 2004. *LCNS*; 3080; 244-260. Springer.
- [7] Ekanayake CC, Dumas M, Garcia-Banuelos L, Rosa ML. Slice, Mine and Dice: Complexity-Aware Automated Discovery of Business Process Models. BPM 2013. *LCNS*; 8094; 49-64. Springer.
- [8] Mamaliga T. Master Thesis: Realizing a Process Cube Allowing for the Comparison of Event Data. 2013. TU/e.
- [9] Sutrisnowati RA, Pulshashi IR, Bae H, Jeon D, Sul S-O. Time-gap analysis considering multi-dimensional perspective for process performance evaluation. *ICIC Express Letters Part B: Applications* 2015; 9:1435-1440.
- [10] Weerd J.D., Broucke S.K.L.M., Caron F. Bidimensional Process Discovery for Mining BPMN Models. DeMiMoP 2014.
- [11] Li J, Liu D, Yang B. Process Mining: Extending alpha-algorithm to mine duplicate tasks in process logs. 2007. *LNCS*; 4537; 396-407. Springer.
- [12] Camunda. Open source platform from workflow and business process automation. <https://camunda.org/bpmn/examples/>. Accessed August 1, 2015
- [13] Carmona J, Cortadella J., Kishinevsky M. A region-based algorithm for discovering petri nets from event logs. BPM 2008. *LCNS*; 5240; 358-373. Springer.
- [14] Medeiros AKA. PhD Thesis: Genetic Process Mining. 2006. TU/e.
- [15] Bergenthum R, Desel J, Lorenz R, Mauser S. Process Mining Based on Regions of Languages. BPM 2007. *LCNS*; 4714: 375-383. Springer.
- [16] Vazquez-Barreiros B, Mucientes M, Lama M. Mining Duplicate Tasks from Discovered Processes. ATAED 2015.
- [17] Process Mining Group Math&CS department, Eindhoven University of Technology ProM; Process mining research tools application. <http://www.processmining.org/prom/start>. Accessed August 1, 2015.
- [18] Sutrisnowati RA, Bae H, Song M. Bayesian network construction from event log for lateness analysis in port logistics. Computers and Industrial Engineering, (in press) (2014)